

# Elsevier Electronic Warehouse

*Architectuurbeschrijving*



# Elsevier Electronic Warehouse

## Architectuurbeschrijving

Inleiding.....	3
Elsevier.....	3
Markt.....	4
Kernwaarden .....	4
Cultuur.....	4
Organisatie .....	4
Global Production.....	5
Doelstellingen van Global Production .....	5
Strategische Keuzes .....	6
De Stakeholders en hun belangen.....	7
De Viewpoints van de Architectuur.....	8
Functioneel Viewpoint.....	8
Ontwerp Viewpoint.....	10
Software Technologie Viewpoint.....	12
Hardware Viewpoint .....	13
Tot Slot .....	15

# Elsevier Electronic Warehouse

## Architectuurbeschrijving

### INLEIDING

Dit document beschrijft de architectuur van het Electronic Warehouse bij Elsevier. In tegenstelling tot wat de naam doet vermoeden, is dit géén Data Warehouse maar een dataverwerkend productiesysteem met een database omvang van 10 Terabyte.

De bedrijfstaal bij Elsevier is Engels, maar ten behoeve van het Nederlands Kampioenschap ICT Architectuur 2005 is dit document in het Nederlands geschreven. Desondanks ontkomen we er niet aan om bepaalde termen in het Engels te noemen: een Nederlandse vertaling zou slechts tot verwarring leiden.

In deze architectuurbeschrijving geven we eerst een beschrijving van het bedrijf Elsevier. De doelstellingen en waarden van het bedrijf vormen de context waarin de architectuur tot stand gekomen is.

### ELSEVIER

**Elsevier heeft 60% van de markt voor wetenschappelijke tijdschriften.**

Elsevier (voorheen Elsevier Science) is 's- werelds grootste leverancier van wetenschappelijke, technische en medische informatie. Elsevier publiceert meer dan 20.000 producten en diensten, zoals wetenschappelijke tijdschriften, boeken, elektronische producten, diensten, databases en portals voor wetenschappelijke, technische en medische communities.

Elsevier heeft haar hoofdkantoor in Amsterdam. Het is een van de 4 divisies van de Reed Elsevier Group plc, een toonaangevende uitgever en informatieleverancier.

Elsevier heeft zich de afgelopen jaren getransformeerd van een traditionele uitgever naar een dienstverlener die informatie beschikbaar maakt in de breedste zin des woords. Dit blijkt zeer duidelijk uit "The Elsevier Mission": *"Elsevier is an integral partner with the Scientific, Technical and Health Communities, delivering superior information products and services that foster communication, build insights, and enable individual and collective advancement in scientific research and health care."*

Elsevier heeft de afgelopen jaren bijna alle daadwerkelijke productietaken uitbesteed. In essentie zorgt het bedrijf nu voor het werven van informatie en het verkopen ervan. De productie bestaat uit het coördineren van de processen tussen de verschillende onderaannemers. Als zodanig heeft Elsevier een regisseursrol binnen het uitgeefproces.

## Markt

Elsevier heeft jarenlang geopereerd in een stabiele, rustige markt. De laatste jaren is dat echter langzaam veranderd. De markt voor wetenschappelijke uitgaven is in beweging: er vindt een verschuiving plaats van papieren publicaties naar elektronische publicaties en er zijn initiatieven ontstaan waar wetenschappelijke publicaties gratis toegankelijk zijn. Tot nu toe blijft Elsevier echter een dominante rol spelen binnen de markt, als gevolg van:

- **Omvang:** Elsevier geeft een kleine 2000 wetenschappelijke tijdschriften (Journals) uit, waarin jaarlijks 250.000 artikelen worden gepubliceerd, tezamen zo'n 2.500.000 pagina's. Hierdoor kunnen er eenvoudig schaalvoordelen behaald worden. Klanten profiteren van het one-stop-shopping principe.
- **Kwaliteit:** De wetenschappelijke uitgaven van Elsevier zijn befaamd om hun constante hoge kwaliteit, zowel qua inhoud als qua presentatie.
- **Toegang:** Elsevier beheert een groot aantal websites waarop de content te bekijken is. Sommige bieden de volledige content aan, andere zijn gericht op disciplines en specialismen. Hierdoor kunnen klanten snel en gemakkelijk de voor hen relevante informatie te vinden. Een voorbeeld hiervan is de website [thelancet.com](http://thelancet.com) van het bekende Journal "The Lancet".

## Kernwaarden

Om de missie van Elsevier kracht bij te zetten, worden er een aantal kernwaarden gehanteerd:

- *Customer Focus*
- *Valuing our People*
- *Passion for Winning*
- *Innovation*
- *Boundarylessness*

## Cultuur

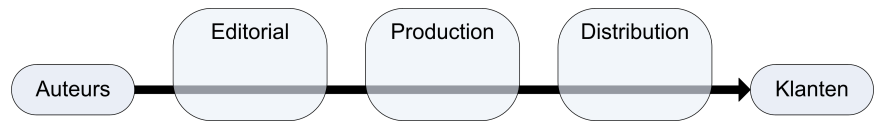
De cultuur van Elsevier wordt gekenmerkt door een hoge mate van gedrevenheid en een constante focus op kwaliteit en doorlooptijd. Daarnaast is er een sterke drang om zich te onderscheiden van de concurrentie door superieure producten en diensten te leveren.

## Organisatie

Elsevier heeft een tweetal marktgerichte divisies, te weten Health Sciences en Science & Technology. Deze divisies richten zich met name op de acquisitie en verkoop van content. Daarnaast is er de divisie Operations waarbinnen het daadwerkelijke uitgeefproces plaatsvindt.

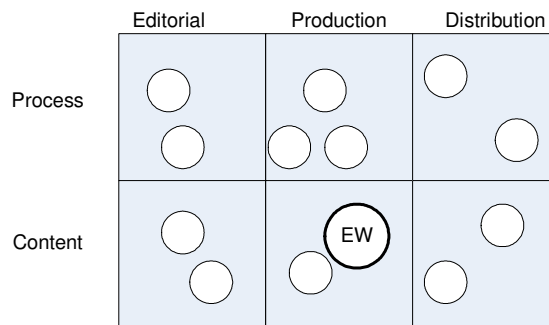
## GLOBAL PRODUCTION

Het uitgeefproces kan grofweg in drie stappen worden weergegeven, zoals blijkt uit onderstaande afbeelding.



Binnen het Editorial proces worden artikelen gereviewed en vinden er vaak verbeterlagen plaats door de auteur. Feitelijk zorgt dit proces voor de inhoudelijke kwaliteitsgarantie. Binnen het Production proces wordt het artikel van het formaat waarin de auteur het aanlevert geconverteerd, gestructureerd, opgemaakt en voorzien van verwijzingen naar referenties. Tenslotte vindt binnen het Distribution proces de fysieke distributie en de plaatsing van het artikel op de diverse websites van Elsevier plaats.

De verzameling van applicaties die bij deze processen gebruikt worden, kan ingedeeld worden volgens onderstaand diagram.



Horizontaal zien we de 3 stadia van het uitgeefproces. De verticale indeling geeft weer dat ieder systeem gericht is óf op het proces, óf op de content. Een indeling op deze wijze, met heldere interfaces tussen de systemen, geeft de mogelijkheid om individuele systemen te verbeteren c.q. te vervangen in beheersbare projecten.

Deze architectuurbeschrijving beperkt zich tot het Electronic Warehouse (EW): het **Content** systeem voor **Production**.

### Doelstellingen van Global Production

Het Production stadium van het uitgeefproces wordt uitgevoerd door de afdeling Global Production. Deze afdeling wordt centraal aangestuurd en maakt wereldwijd gebruik van een gestandaardiseerd proces en gestandaardiseerde systemen.

De doelstellingen van Global Production ondersteunen de Corporate doelstellingen:

- **Doorlooptijd van auteur naar publicatie:** voor onderzoekers en wetenschappers is het van cruciaal belang dat vindingen snel

gepubliceerd worden. De eerste die een vinding publiceert is de winnaar; er is geen tweede prijs.

- **Kwaliteit van de publicaties:** De reputatie van een Journal bepaalt voor een belangrijk deel de betrouwbaarheid van de inhoud. Wetenschappers kiezen bij voorkeur een gezaghebbend Journal voor hun publicatie; bibliotheken nemen dit eveneens in overweging bij de samenstelling van hun abonnementen.
- **Kostenbeheersing:** Met toenemende concurrentie van andere uitgevers en initiatieven in de bestaande markt is kostenbeheersing een ongoing onderwerp. Voor de systemen kan dit worden vertaald naar lage beheerkosten.
- **Flexibiliteit:** Het innoverend vermogen van Elsevier wordt in belangrijke mate bepaald door de flexibiliteit van de processen en de ondersteunende systemen.

### Strategische Keuzes

Om deze doelstellingen te realiseren, heeft Elsevier de volgende keuzes gemaakt:

- Uitvoerende processen worden zoveel mogelijk uitbesteed. Dit zorgt voor flexibiliteit en kostenbeheersing
- Regie van het Production proces is de kerntaak. Proceskennis blijft in huis en doorlooptijd kan beïnvloed worden

Het resultaat hiervan is dat Elsevier de regiefunctie vervult binnen een (wereldwijde) netwerkorganisatie. Dit zorgt voor flexibiliteit en kostenbeheersing in de processen, maar verhoogt de noodzaak tot constante aandacht voor kwaliteit.

Procesbeheersing en –verbetering wordt dan ook gezien als de belangrijkste aandachtspunten bij Global Production. Door efficiënte inrichting van de processen kan de kwaliteit geborgen worden en de kosten beperkt. Continue procesverbetering blijft nodig om de dominante positie te behouden. De processen zijn daarbij leidend voor de systemen. Zij moeten voldoende flexibel zijn om deze toekomstige veranderingen aan te kunnen.

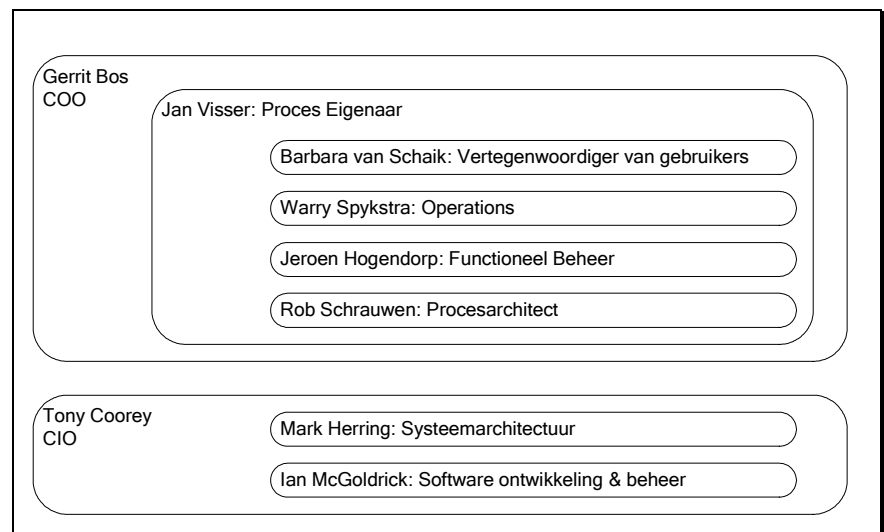
Om bovenstaande redenen is bewust gekozen voor maatwerkapplicaties in plaats van standaardapplicaties die vervolgens aangepast worden.

“Margin improvement through continuing process efficiency.”

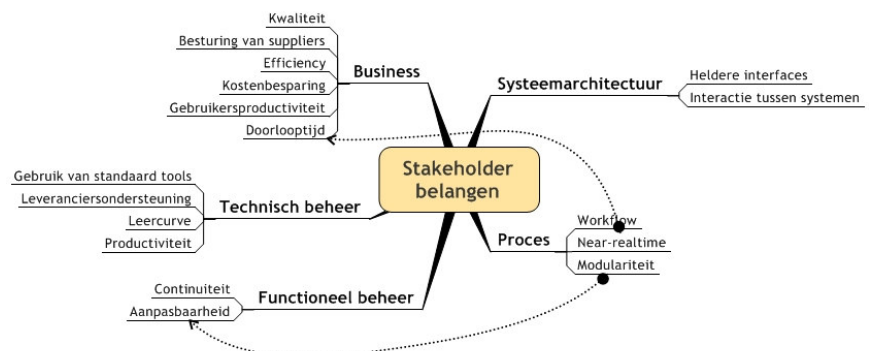
—Sir Crispin Davis,  
CEO, Reed Elsevier

## De Stakeholders en hun belangen

De stakeholders zijn in onderstaand diagram weergegeven. Dit is geen organigram; de kaders geven de verantwoordelijkheden aan.



De *stakeholder belangen* zijn geïnventariseerd en gegroepeerd in onderstaand mindmap diagram.



Deze belangen zijn de drivers voor de totstandkoming van de architectuur. In het volgend hoofdstuk, waarin de architectuur wordt beschreven, zijn de belangen terug te vinden.

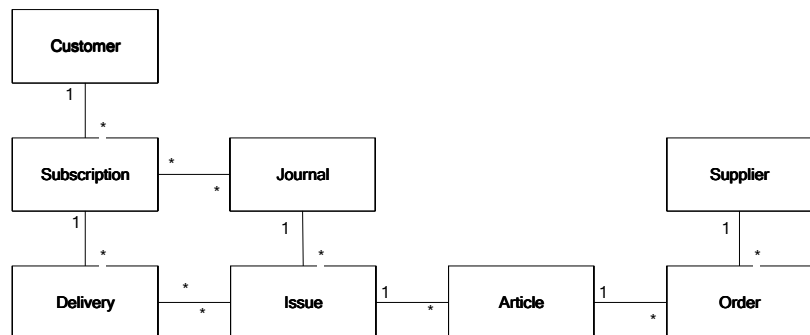
## DE VIEWPOINTS VAN DE ARCHITECTUUR

Een architectuur kan beschreven worden vanuit verschillende gezichtpunten, in het jargon ook wel *viewpoint* genoemd. In dit hoofdstuk worden de viewpoints van de EW architectuur beschreven, te weten:

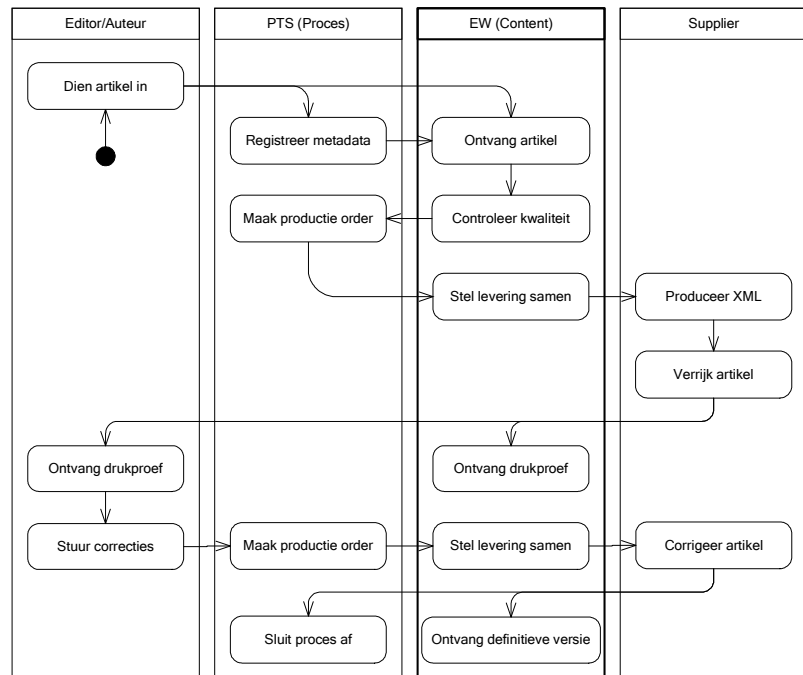
- **Functioneel:** Dit geeft een overzicht van de functionaliteit van het systeem. Doelgroepen zijn het management en de eindgebruikers.
- **Ontwerp:** Dit geeft een overzicht van de logische componenten en hun samenhang. Het ontwerp vormt de brug tussen de functionaliteit en de technologie.
- **Software Technologie:** De toegepaste software producten worden hier beschreven. Dit is met name interessant voor ontwikkelaars en (technisch) beheerders.
- **Hardware:** De hardware componenten en hun samenhang worden in het hardware viewpoint beschreven. Dit viewpoint is bedoeld voor de systeemarchitect en technisch beheerders.

### Functioneel Viewpoint

Onderstaand class diagram geeft een sterk vereenvoudigd overzicht van de classes die een rol spelen bij de productiesystemen. Centraal hierbij is een *article*. Een article zal meerdere keren naar een *supplier* toegestuurd worden via een *order*. De order zal aangeven wat de supplier geacht wordt te doen met het artikel (bijvoorbeeld correcties aanbrenge). Nadat een article gereed is, zal het verschijnen in een bepaald *Issue* van een *Journal*. *Customers* hebben *Subscriptions* op Journals en ontvangen de issues in een periodieke *Delivery*.



Het pad dat een artikel door het productieproces aflegt kunnen we in een activity model volgen:



De actoren in dit proces zijn:

- **Editor/Auteur:** dient het artikel in en reviewt het resultaat van het productieproces. Deze actoren zijn géén medewerkers van Elsevier. Dit aspect onderscheidt een wetenschappelijke uitgeverij van andere uitgeverijen.
- **PTS:** Het *Production Tracking System* bestuurt het proces.
- **EW:** Het *Electronic Warehouse* bewerkt en controleert de artikelen.
- **Supplier:** De supplier (toeleverancier) voert opdrachten uit volgens van tevoren overeengekomen kwaliteitscriteria.

We zien hier de prominente rol van de supplier: hij maakt integraal deel uit van de keten. De uitwisseling van content en metadata tussen Elsevier en de suppliers is in hoge mate gestandaardiseerd. Hierdoor kan Elsevier flexibel met de suppliers omgaan. Zo zijn suppliers wereldwijd verspreid, van Litouwen tot de Filippijnen. Niet expliciet getekend zijn de rigide controles op de toelieferingen. Deze betreffen zowel de kwaliteit van de levering als de levertijd na het uitsturen van een order.

Dagelijks worden er zo'n 1000 nieuwe artikelen van auteurs ontvangen en worden er zo'n 2000 artikelen van suppliers ontvangen. Gemiddeld worden er drie versies van een artikel ontvangen voordat de uiteindelijke versie gereed is.

## Ontwerp Viewpoint

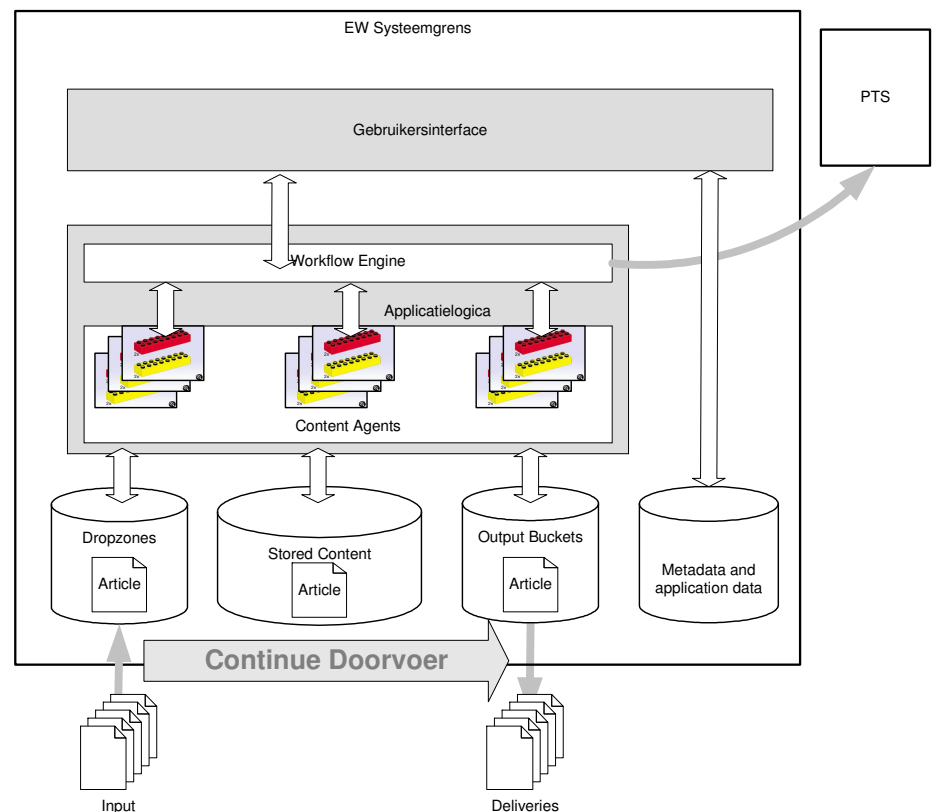
Dit viewpoint beschrijft het (logisch) ontwerp van het systeem. Het systeem is zodanig ontworpen, zodat het een specifieke bijdrage levert aan de doelstellingen van de afdeling Global Production:

1. Vergroten van de snelheid en de efficiency van het productieproces. De doorlooptijd van auteur naar publicatie moet zo kort mogelijk zijn, met een minimum aan handmatige handelingen.
2. Verhogen van de flexibiliteit bij het produceren en distribueren van elektronische en papieren content. Het systeem dient zowel het proces voor boeken als het proces voor Journals te ondersteunen.
3. Kostenreductie door centrale opslag van content.

Aangezien het EW ook als vervanging van een ander systeem functioneert, moest het tevens in een aantal tekortkomingen van het oude systeem voorzien:

4. Verhogen van de capaciteit, snelheid en beschikbaarheid;
5. Betere ondersteuning voor gebruikers bij het productieproces;
6. Ondersteuning voor het gelijktijdig verwerken van input en produceren van output (bij het oude systeem was batchgewijs input óf output mogelijk).

Om aan de punten 1 en 2 tegemoet te komen is er gebruik gemaakt van een modulaire drielaagstructuur.



De gebruikers interface bevat de schermen waarmee de interactie tussen gebruiker en systeem plaatsvindt. In de applicatie logica laag bevinden zich de workflow definities die het productieproces aansturen. Tevens bevinden zich in deze laag de content agents, die diverse bewerkingen en controles uitvoeren op de content en aangestuurd worden door de workflow engine. In de data laag bevindt zich de database waarin de metadata en content wordt opgeslagen. Tevens wordt hier de inkomende en uitgaande content ontvangen c.q. verstuurd, om zodoende punt 3 te verwezenlijken.

### **Bucket**

Een bijzonder onderdeel van het ontwerp is de zogenaamde Output Bucket, een essentieel onderdeel om de punten 4 en 6 te implementeren. Zodra er content wordt ontvangen, wordt er onmiddellijk nadat de kwaliteitscontroles met succes zijn afgerond gecontroleerd of en welke klanten geabonneerd zijn op deze content. Als er relevante abonnementen gevonden worden, wordt de content toegevoegd aan een zogenaamde bucket. De content druppelt als het ware in de bucket. In de buckets wordt alle voor een klant relevante content tijdelijk opgeslagen totdat de daadwerkelijke delivery plaatsvindt. De frequentie van deliveries aan klanten varieert van onmiddellijk tot tweewekelijks.

### **Upload/Download Tool**

Centrale opslag van de content (punt 3) maakt het mogelijk om de ondersteuning van gebruikers (punt 5) verder te verbeteren. Door een eenvoudige gebruikersinterface te maken waarmee elke gebruiker content in het EW kan plaatsen en het er ook weer uit kan halen, is het mogelijk geworden om naast de verschillende versies van het artikel ook allerlei andere gerelateerde documenten op te slaan, zoals correspondentie met de auteur (schriftelijk op per e-mail), aantekeningen en opmerkingen van productiemedewerkers enz. Deze Upload/Download Tool is wereldwijd beschikbaar voor alle medewerkers van Elsevier. Elk artikel heeft op deze manier een eigen “dossier” waarin alle relevante informatie terug te vinden. Dit zorgt voor een aanzienlijke verbetering in de doorlooptijd en kwaliteit van het productieproces.

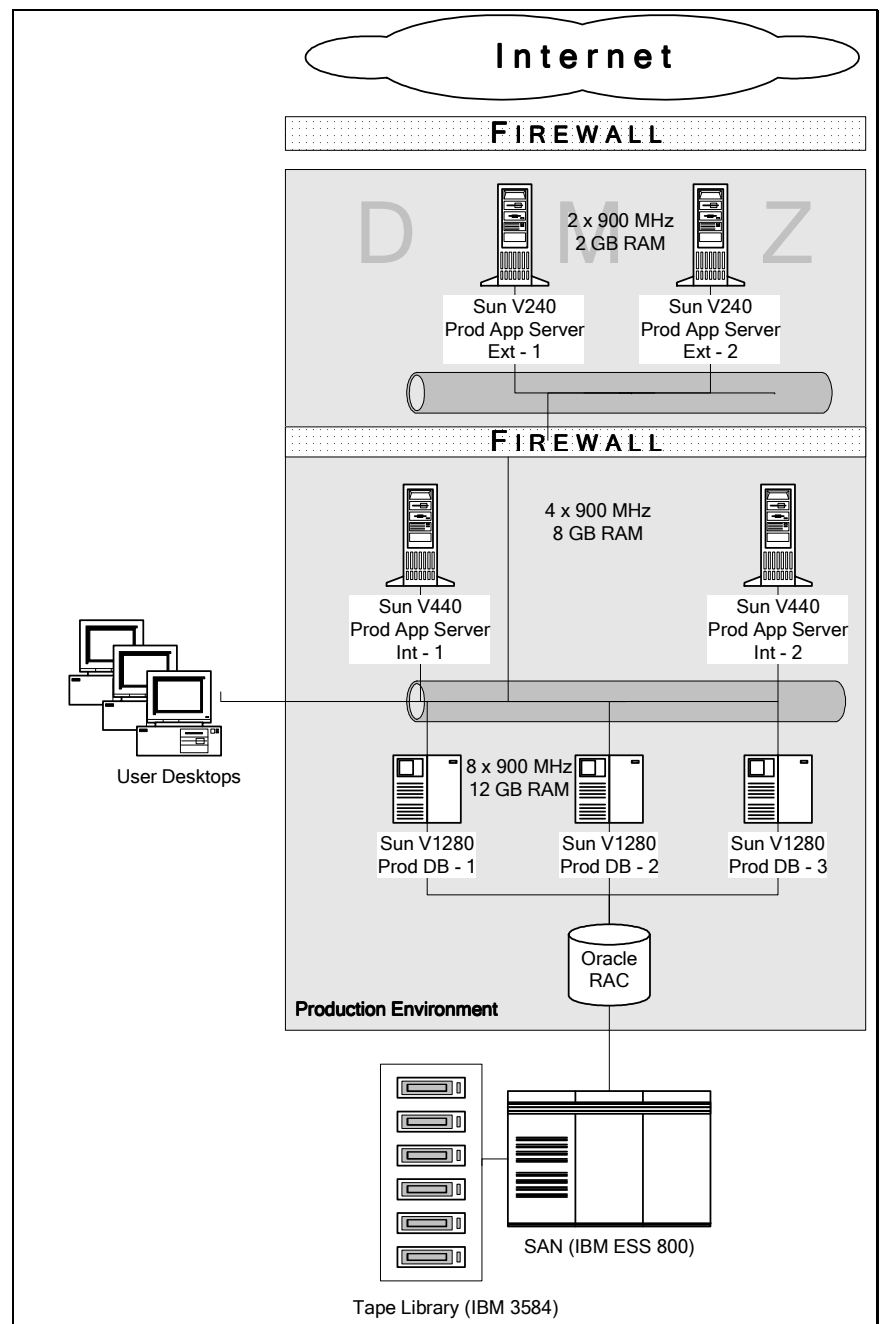
## Software Technologie Viewpoint

Voor de ontwikkelaars van de software is het van belang met welke tools zij het systeem bouwen en onderhouden.

- **Oracle Database:** Alle content wordt op één plaats opgeslagen zodat de content (artikelen) en de bijbehorende metadata altijd consistent zijn. De omvang van de database bedraagt 10 Terabyte. De Oracle database biedt standaard faciliteiten om deze enorme hoeveelheid gegevens om te gaan. De database is geconfigureerd als een **Real Application Cluster (RAC)**, wat betekent dat de database over resources van meerdere servers kan beschikken. Dit betekent dat de database verwerkingscapaciteit eenvoudig uit te breiden is, maar ook dat de database beschikbaar blijft wanneer een server uitvalt.
- **Oracle Content Management SDK:** Deze software development kit (SDK) implementeert een toegangslaag om de database. Hiermee kunnen bestanden via diverse protocollen (zoals FTP en email) rechtstreeks van of naar de database gestuurd worden. Dit verhoogt de beheerbaarheid (geen tussentijdse opslag) en is door de gelaagde structuur goed schaalbaar.
- **Java:** De programmatuur die de content bewerkt is geschreven in Java. Java is bij uitstek geschikt voor XML verwerking. Dit deel van het systeem is object-georiënteerd opgezet, met de gebruikelijke voordelen hiervan. De Java programmatuur runt in de **Oracle Application Server** en is hiermee met de **Enterprise Manager** te beheren, samen met de database.
- **Oracle Workflow:** Alle workflow processen zijn hiermee gemodelleerd en gebouwd. De hechte integratie met de database, schaalbaarheid en snelheid waren redenen om deze component te gebruiken. Bovendien was Oracle Workflow reeds met veel succes in het PTS systeem in gebruik. Hierdoor was het product reeds bekend bij de technisch beheerders.
- **Oracle Forms:** De gebruikers interface is met Oracle Forms gebouwd. Hiermee is een rijke gebruikersinterface gerealiseerd waarmee de gebruikers bovendien al bekend waren via andere systemen. Oracle Forms draait in de browser van de gebruiker zodat beheerkosten minimaal zijn. De beheerorganisatie was reeds bekend met Oracle Forms.

## Hardware Viewpoint

De hardwareconfiguratie is ontworpen om te voldoen aan de eisen van hoge beschikbaarheid en schaalbaarheid. We zien daarom dat vrijwel alle componenten dubbel uitgevoerd zijn. In plaats van één grote database server is gekozen voor 3 kleinere database servers. Dit leverde een significante besparing op de hardwarekosten. Beheerkosten van de extra servers zijn slechts marginaal hoger door gebruik te maken van de meegeleverde Oracle beheertools.



De configuratie bestaat uit drie lagen:

- Externe applicatie servers  
Deze servers bevinden zich in de DeMilitarized Zone (DMZ) en worden uitsluitend gebruikt voor toegang tot de applicatie door suppliers en klanten.
- Interne applicatie servers  
Deze servers bevinden zich binnen het lokale netwerk van Elsevier en worden gebruikt door alle eindgebruikers.
- Database servers  
Voor de database wordt gebruik gemaakt van 3 identieke Sun servers die door middel van een high speed interconnect aan elkaar verbonden zijn. Voor opslag wordt gebruik gemaakt van een Storage Area Network (SAN) van IBM met een capaciteit van 20 TB.

#### **Netwerk**

Het Electronic Warehouse verwerkt dagelijks zo'n 16 GB aan inkomend verkeer, en 80 GB aan uitgaand verkeer. Dit heeft zijn weerslag op de benodigde netwerkcapaciteit op het wereldwijde Elsevier netwerk.

#### **Storage en Backup**

De omvang van de database is 10 Terabyte. Dat dit een zeer respectabele omvang is blijkt uit het feit dat deze database de 3e plaats inneemt op de ranglijst van 's werelds grootste OLTP databases<sup>1</sup>. Gangbare backup oplossingen zouden te traag zijn om deze hoeveelheid binnen een redelijke tijd op tape te zetten, terwijl ze bovendien het systeem belasten tijdens de uitvoering van de backup. Een zgn. LAN-free backup faciliteit maakt het mogelijk om de backups op een beheerste manier te maken terwijl het systeem volledig beschikbaar blijft gedurende de backups.

---

<sup>1</sup> Zie het onderzoek van Winter Corporation op <http://www.wintercorp.com/>

## TOT SLOT

Met het Electronic Warehouse zijn de doelstellingen van Elsevier Global Production ondersteund, te weten:

- Doorlooptijd van auteur naar publicatie
- Kwaliteit van de publicaties
- Kostenbeheersing
- Flexibiliteit

De gekozen architectuur maakte het mede mogelijk om het EW te implementeren in een project dat op tijd en binnen het budget gereed was.

De resultaten zijn niet alleen zichtbaar binnen Elsevier, maar ook daarbuiten. De snellere doorlooptijd is merkbaar voor zowel auteurs als klanten. Publicaties zijn op verscheidene media en vormen beschikbaar, zoals bijvoorbeeld volledige publicaties versus samenvattingen, papier versus elektronisch en PDF versus HTML.



**ELSEVIER**

Elsevier Electronic Warehouse: Architectuurbeschrijving

September 2005

Auteurs: Rob Swinkels, Michael Magan (Oracle)

Met bijdragen van: Jeroen Hogendorp, Rob Schrauwen, Warry Spykstra (Elsevier)